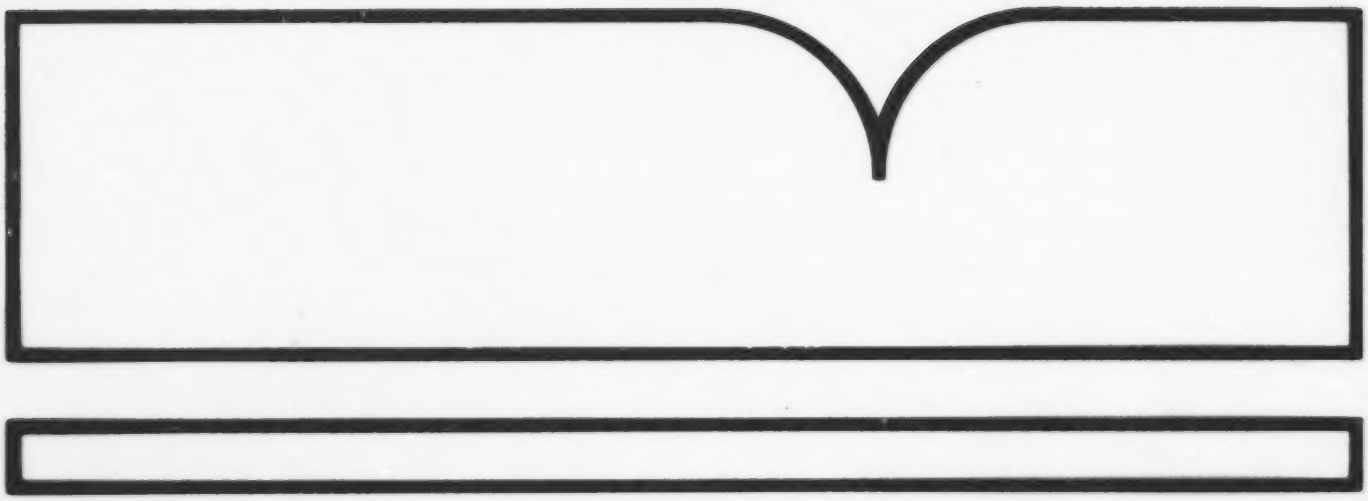
PB90-216102

Evaluation of Long-Term Sulfur Deposition Models

(U.S.) Environmental Protection Agency, Research Triangle Park, NC

1989



U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA/600/J-89/345		PB 90 216102/AS

4. TITLE AND SUBTITLE	5. REPORT DATE
THE EVALUATION OF LONG-TERM SULFUR DEPOSITION MODELS	6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
T.L. Clark,[1] R.L. Dennis[1] E.C. Voldner,[2] M.P. Olson,[2] S.K. Seilkop,[3] M. Alvo[4]	

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
1)ASRL/MAD, RTP, NC 27709 2)AES, Downsview, Ont. 3)Anal. Sci. Inc., RTP, NC 4)University of Ottawa, ONT	N104/C/05/05-3166 (FY-88) 11. CONTRACT/GRANT NO.

12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Atmospheric Sciences Research Laboratory-RTP, NC Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, NC 27711	14. SPONSORING AGENCY CODE EPA/600/09

15. SUPPLEMENTARY NOTES

16. ABSTRACT

The International Sulfur Deposition Model Evaluation (ISDME) project, jointly conducted by the U.S. Environmental Protection Agency and Atmospheric Environment Service of Environment Canada, assessed the performance of eleven linear chemistry atmospheric models in predicting amounts of sulfur wet deposition. Standardized model input data sets were distributed to the participating modelers, who later submitted seasonal and annual 1980 model predictions of dry/wet deposition and air concentrations of sulfur dioxide and sulfate at up to 65 sites across eastern North America. The models were evaluated in an operational mode using new, more rigorous approaches, as well as the more conventional distribution statistics recommended by the American Meteorological Society. The new approaches focused on the ability of the models to replicate features of the spatial patterns of sulfur wet deposition, as determined by an interpolation technique known as kriging. This technique quantified the uncertainties in the observations which were used in the evaluation process to identify areas where interpolated predictions were statistically significantly different from the interpolated observations. To supplement the evaluation, predictions of dry deposition amounts and air concentrations of each model were intercompared to identify apparent peculiarities. Finally, a scoring system based on criteria for six model performance measures was devised to compare seasonal, annual and overall performances of the models. Three clusters of models, each with similar overall scores, were identified.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL INFORMATION SERVICE
SPRINGFIELD, VA. 22161

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	23
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

PB90-216102

Atmospheric Environment Vol. 23, No. 10, pp. 2267–2288, 1989.
Printed in Great Britain.

NTIS is authorized to reproduce and sell this report. Permission for further reproduction must be obtained from the copyright owner.

Pergamon Press plc

EPA/600/J-89/345

THE EVALUATION OF LONG-TERM SULFUR DEPOSITION MODELS

TERRY L. CLARK,* EVA C. VOLDNER,† ROBIN L. DENNIS,* STEVEN K. SEILKOP,‡ MAYER ALVO§ and MARVIN P. OLSON†

*U.S. Department of Commerce, on assignment to U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A., †Atmospheric Environment Service, Downsview, Ontario, Canada M3H 5T4, ‡Analytical Sciences Incorporated, Durham, NC 27713, U.S.A. and §Department of Mathematics, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5

(*First received* 18 *July* 1988 *and in final form* 30 *January* 1989)

Abstract—The International Sulfur Deposition Model Evaluation (ISDME) assessed the performance of 11 regional, long-term deposition models in predicting amounts of sulfur wet deposition. With few exceptions, each model predicted air concentrations and dry/wet deposition amounts of SO_2 and SO_4^{2-} at up to 66 sites across eastern North America for each season of 1980. Unlike its predecessors, this evaluation focused on the ability of the models to replicate, within the uncertainty limits of the data, the magnitude and position of the seasonal spatial patterns of S wet deposition. Both the spatial patterns and the uncertainties arising from measurement and interpolation errors were determined by a simple Kriging technique. Model performance measures included the per cent of subregions where interpolated predictions and observations were significantly different and differences between the magnitudes and locations of maxima, centroids and orientations of the major axes.

When data uncertainty limits are considered, the seasonal predictions of one Lagrangian and two statistical models were insignificantly different from the seasonal observations (at approximately the 95% confidence level) across at least 85% of the evaluation region. Three other models performed relatively well for two or three seasons, while the performance of three others was somewhat erratic. The final two models significantly underpredicted for every season across at least 40% of the evaluation region.

In spring, all of the models correctly predicted the location of the observed maximum amount of S wet deposition. For the other three seasons the models predicted the locations within 170–350 km. Except for summer, most models predicted the magnitude of the observed maximum within the data uncertainty limits.

Key word index: Model evaluation, acid deposition, sulfur deposition modeling, sulfur emissions, Kriging.

1. INTRODUCTION

Regional long-term sulfur deposition models have been developed to calculate mean air concentrations of SO_2 and SO_4^{2-} and amounts of S wet and dry deposition for monthly, seasonal, and annual periods across Europe and North America. The results of some of these models have been used to estimate the relationships between pollutant source regions and receptor regions on regional and international scales. These relationships, expressed as source–receptor matrices, have been applied to estimate the reduction of sulfur deposition amounts (MOI, 1982) and the improvement to visibility as a consequence of nonuniform reductions of sulfur emissions (U.S. Environmental Protection Agency, 1987). In addition, some of those models have been applied to construct spatial patterns of sulfur dry deposition for use in lake acidification studies (Clark *et al.*, 1989).

For several reasons it is desirable to evaluate these models using a comprehensive set of data (i.e. a data base consisting of values of all parameters calculated by the model). First, model evaluations are essential steps in the model development process. By comparing model calculations (hereafter referred to as predictions) and measurements (referred to as observations), a modeler is able to identify deficiencies in the model algorithms and modify the parameterizations to improve the performance of the model. Secondly, from the model evaluations the model-user community is provided information regarding the uncertainty of the predictions and the credibility of the model. Finally, since the variables predicted by these models are interrelated, a comprehensive data set is essential to determine the causes of the differences between predictions and observations.

Since comprehensive data sets were unavailable, the objectives of previous model evaluations were restricted to model performance assessment. The European evaluations assessed model performance on the basis of comparisons of monthly or annual mean predictions and observations of concentrations in air and precipitation at as many as 50 sites (Eliassen, 1978; Norwegian Meteorological Institute, 1984; Klug and Lupkes, 1985; Lehmhaus *et al.*, 1986). Similarly, the North American evaluations assessed model performance on the basis of comparisons of monthly or annual mean predictions and observations at as many as 54 sites for air concentrations, but only as many as 15 sites for wet deposition (McNaughton, 1980; Voldner *et al.*, 1981; Bhumralkar *et al.*, 1981; Shannon *et al.*, 1982; Voldner and Shannon, 1983; Stewart *et al.*,

1983a,b; Schiermeier and Misra, 1983; Ruff *et al.*, 1985).

Until the middle of this decade the daily SO_2 and SO_4^{2-} air concentrations measured during the 1977 and 1978 SUlfur Regional Experiment (SURE) (Perhac, 1978) provided the only North American evaluation data base suitable for statistically meaningful evaluations of regional, long-term deposition models. During this period neither dry deposition measurements nor sufficient S wet deposition measurements were available for a statistically meaningful evaluation. By 1980, however, the number of precipitation chemistry sites in North America significantly increased as a result of the expansion of existing networks and the establishment of additional networks. The significant increase in the number of wet deposition monitoring sites now made it feasible to evaluate these long-term models for S wet deposition amounts.

Recognizing the need to evaluate these regional models on a common basis and to determine whether they satisfactorily replicate S wet deposition patterns, the Office of Research and Development of the U.S. Environmental Protection Agency and the Atmospheric Environment Service of Environment Canada conducted the International Sulfur Deposition Model Evaluation (ISDME). This study evaluated 11 regional models using seasonal S wet deposition amounts calculated from screened 1980 precipitation chemistry data. Although the volume of precipitation chemistry data was increasing for subsequent years, 1980 was selected for the evaluation since it was the only year that a detailed emissions inventory was available.

2. ISDME OBJECTIVES

It is recognized that a comprehensive model evaluation should not only investigate whether a model produces the *right results*, but it also should address the question, did the model produce the right results for the *right reasons*? This question needs to be answered to establish the level of confidence one would have that a model that produced the right results for one scenario would also produce the right results for another scenario.

Since simultaneous measurements of concentrations in air and precipitation were available for only a few Canadian sites, a comprehensive model evaluation was not possible. Therefore, it was impossible to substantiate whether those models producing the right results did so for the right reasons. However, these air concentration data were useful in identifying inconsistencies in model results. In addition, the lack of a credible means of empirically defining actual source–receptor relationships precluded an evaluation of the modeled relationships. However, with the available precipitation chemistry data, the performance of the models in predicting S wet deposition is assessed by this study; detailed explanations of model behavior are not addressed.

The specific objectives of the ISDME were as follows.

(1) Evaluate operational S deposition models for each season for an entire year.
(2) Identify which models best replicate the observed spatial and temporal patterns of sulfur wet deposition within data uncertainty limits.
(3) Delineate for each model, areas and periods of good and poor agreement with the observed sulfur wet deposition amounts with considerations for data uncertainties.
(4) Compare predictions of air concentrations and dry deposition amounts.
(5) Investigate the effects on model performance from the uncertainties of processed precipitation data grids and the observed S wet deposition amounts.
(6) Attempt to explain the model performances.

Previous evaluations of S deposition models relied primarily on the model performance measures recommended by the American Meteorological Society (AMS) (Fox, 1981). These model evaluations focused almost entirely on the comparison of statistical distributions of observations and predictions at measurement sites (e.g. means, biases, standard deviations). Spatial analyses were generally restricted to smoothed analyzed maps of data, location of maxima, spatial correlations and cross-sectional plots.

The AMS model evaluation recommendations and the spatial analyses used in previous model evaluations alone are not sufficient to address the ISDME objectives. Consequently, the ISDME required the development of a new approach with a greatly increased emphasis on spatial analysis. The ISDME approach used distributional statistics from the more traditional approach plus measures relating to a robust spatial analysis with an emphasis on the latter.

In achieving the ISDME objectives, the precipitation chemistry measurements were first screened for completeness; monitoring sites were screened for representativeness of the regional scale (Voldner *et al.*, 1984). In addition, techniques were explored and methods adapted to spatially analyze predictions and observations of S wet deposition while minimizing interpretation errors, and to quantify uncertainties of the measurements and interpolations (Clark *et al.*, 1987).

For the sake of brevity, only the evaluation results addressing the first three and the last goals are discussed here. Clark *et al.* (1986) discuss the effects of the uncertainties of gridded precipitation data on model performance and Clark *et al.* (1987) present additional details of the ISDME study.

3. MODELS

The regional models evaluated in this study (Table 1) are designed to predict mean air concentrations of SO_2 and SO_4^{2-} and total wet/dry deposition amounts of either S or both SO_2 and SO_4^{2-} for periods of a month to a year. With one exception, these models, whose spatial resolution approximates 100 km, simulate the same fundamental processes; SERTAD does not simulate the dry deposition process. The attributes of these models are summarized in the ISDME Final Report (Clark *et al.*, 1987).

The models basically simulate the fate of pollutants by first emitting SO_2 from virtual point sources (at the center of each model grid cell). The SO_2, as well as the SO_4^{2-} resulting from the transformation of the SO_2, are vertically mixed and horizontally diffused at rates prescribed by the model. During transport, percentages of the pollutant mass are dry deposited and scavenged by precipitation according to rates prescribed by the model and the precipitation data used by the model. A portion of the total mass is transported beyond the continental shores, off the model grid, and no longer considered. Some models impose a background level of S wet deposition attributed to biogenic and distant anthropogenic sources.

Although they simulate essentially the same processes, the modeling approaches can differ. Basically, S deposition models can be classified into four genres according to model approach: statistical, Lagrangian, Eulerian (ordered by increasing levels of sophistication) or hybrid. The fundamental differences of these model genres are discussed by Pasquill and Smith (1983). In simple terms, the four ISDME statistical models generally simulate transport and diffusion via one or more long-term mean, single-level wind roses, that represent speed and direction frequencies for 22.5-degree sectors. As in the other ISDME models, the transformation of SO_2 to SO_4^{2-} is a linear process expressed as a function of time. Wet and dry deposition of pollutant mass is dependent on the mean frequency and duration of wet and dry periods rather than the actual occurrences and rates of precipitation.

Except for ASTRAP and RTM-II, the remaining models are Lagrangian in nature in that simulated processes act on the pollutant mass in discrete puffs. These puffs are transported using objectively analyzed winds at one or more levels. Horizontal diffusion is based on puff travel time and diffusion coefficients. Wet deposition is dependent on the pollutant scavenging rate, the rate and duration of precipitation and pollutant air concentrations, while dry deposition is a function of dry deposition velocity and surface air concentration. At each receptor point or grid cell, mean air concentrations and total deposition amounts are determined by adding the contributions of each puff. Details of the attributes of these models are provided by Clark *et al.* (1987).

Rather than considering discrete puffs, ASTRAP uses ensemble statistics (i.e. mean and standard deviations of puff position) to define transport and horizontal diffusion (Shannon, 1985). Separate sets of ensemble statistics are determined for dry and wet puffs. Both dry and wet deposition functions are similar to those of the Lagrangian models.

Finally, RTM-II, a Lagrangian–Eulerian hybrid model, simulates the processes using Gaussian puffs within the 80-km emission, source square and a finite

Table 1. The 11 long-term sulfur deposition models evaluated in the ISDME

Model name	Model genre
Atmospheric Environment Service Long-Range Transport Model (AES)	Lagrangian
Advanced Statistical Trajectory Regional Air Pollution (ASTRAP) Model	Statistical–Lagrangian
Ontario Ministry of the Environment (MOE) Long-Range Transport Model	Lagrangian
Regional Impacts on Visibility and Acid Deposition (RIVAD) Model	Lagrangian
Regional Lagrangian Model of Air Pollution (RELMAP)	Lagrangian
Regional Transport Model (RTM-II)	Lagrangian–Eulerian
Statistical Estimates of Regional Transport and Acidic Deposition (SERTAD) Model	Statistical
Statistical Model (STATMOD)*	Statistical
Tennessee Valley Authority (TVA)	Statistical
University of Michigan Atmospheric Contributions to Interregional Deposition (UMACID) Model	Lagrangian

*A second version of this model, SIMPMOD, was also applied.

difference scheme for transport and diffusion beyond the emission source square (Stewart *et al.*, 1983a). RTM-II algorithms for dry and wet deposition resemble those of Lagrangian models.

Only MOE and SERTAD superimpose a background contribution to the predictions of S wet deposition. Although both models consider background contributions, the definitions of the background levels for the two models are quite different. The MOE background contribution attempts to account for the effects of the S emissions from biogenic sources inside the model domain, which were not included in the anthropogenic emissions inventory, as well as from sources outside the model domain. A total annual background contribution of 2 kg S ha^{-1}, determined empirically, is added to the predicted amounts across each cell in the model domain. This contribution ranges from 30 to 60% of the observed annual amounts at remote receptors, where background levels are relatively significant, and less than 10% of those at receptors in the high deposition region.

On the other hand, the SERTAD background contribution is comprised of two components. The first component, applicable to all seasons but winter, assumes a uniform seasonal background contribution to the precipitation concentration ranging from 1.7 to 2.6 mg SO_4^{2-} ℓ^{-1} (Concord Scientific Corp., 1985). The effective annual contribution is approximately a factor of 3 greater than the measured precipitation concentrations at remote Canadian monitoring sites (Barrie and Hales, 1984). It is important to note that before this study, these background levels were determined in part by comparing sets of predictions to a set of multiyear observations, which included the ISDME evaluation year. Thus, this model could have an intrinsic advantage over the other models.

The second component assumes that a minimum of 80% of the S emissions from stacks exceeding 250 m is wet deposited across the model domain. For both components, the amount deposited across a receptor grid cell is proportional only to the precipitation amount for that grid cell; source–receptor distances are irrelevant to the amount deposited. Hence, the spatial pattern of SERTAD predictions of S wet deposition would resemble, to some degree, the spatial pattern of precipitation and would be less sensitive than other models to changes in the emission patterns.

Background levels for SERTAD were much higher than those for MOE. For example, based on typical precipitation rates across Ohio during the summer of 1980, the SERTAD background contributions from only the first component are 2.4 kg S ha^{-1}, or about 40% of the observed summer amount.

For the ISDME applications each of these models used the standardized set of seasonal 1980 grids of SO_2 emissions (Fig. 1), which were allocated to three layers (0–200, 201–700 and >700 m) according to effective stack heights (Clark, 1985). The sources of the 1980 emissions data were the Version 2.0 National Acid Precipitation Assessment Program (NAPAP) inventory for the U.S. and the Canadian inventory used in the U.S. Canadian MOI project. In addition, the models used either the 1980 standardized meteorological data set provided by Clark *et al.* (1985) or the modeler's pre-existing data set.

4. EVALUATION DATA

The observed amounts of S wet deposition comprising the ISDME evaluation data base are derived from 1980 measurements of the Air and Precipitation Monitoring Network (APN) and Canadian Network for Sampling Precipitation (CANSAP) in Canada and the Electric Power Research Institute (EPRI) Network, Multistate Atmospheric Power Production Pollution Study (MAP3S) Network and the National Atmospheric Deposition Program (NADP) Network in the U.S. The APN, EPRI and MAP3S networks collect precipitation samples on a daily basis, the NADP network on a weekly basis and the CANSAP network on a monthly basis. The precipitation data were retrieved from the Acid Deposition System (Watson and Olsen, 1984).

Since the methods for quality control and quality assurance of the precipitation chemistry data and the procedure for computing wet deposition amounts are not identical for the aforementioned networks, it was necessary to adopt a unified procedure for screening the precipitation chemistry data and computing deposition amounts. Furthermore, since the level of uncertainty of the seasonal wet deposition amounts calculated at each monitoring site may differ (due to missing or rejected data for instance), a screening procedure was developed to classify the data into three overall data quality levels (Unified Deposition Data Base Committee, 1985). This procedure was based on two factors: site representativeness of the regional scale (to identify sites perceived to be influenced by local sources) and data completeness (to avoid those seasonal amounts that are based on an unacceptable fraction of the sampling period). Site representativeness was based on documented site characteristics and the proximity of the site to major point sources of SO_2 (Unified Deposition Data Base Committee, 1985). Data completeness was based on the percentage of the sampling period when valid data were available, precipitation collection efficiency, and degree of contamination of sample from sea spray. Details are provided in Appendix C of Clark *et al.* (1987).

Seasonal amounts categorized as level 1 passed the most stringent screening criteria, while level 3 amounts failed to satisfy the minimum criteria. Level 2 amounts either were based on less reliable data than level 1 amounts or were based on data collected at a site that could have been influenced by 'local' sources. The ISDME evaluation data set for the U.S. consisted of only data categorized as level 1 or level 2. The same was true for the Canadian data, with a few exceptions; several sites in the level 3 category, strictly on the basis

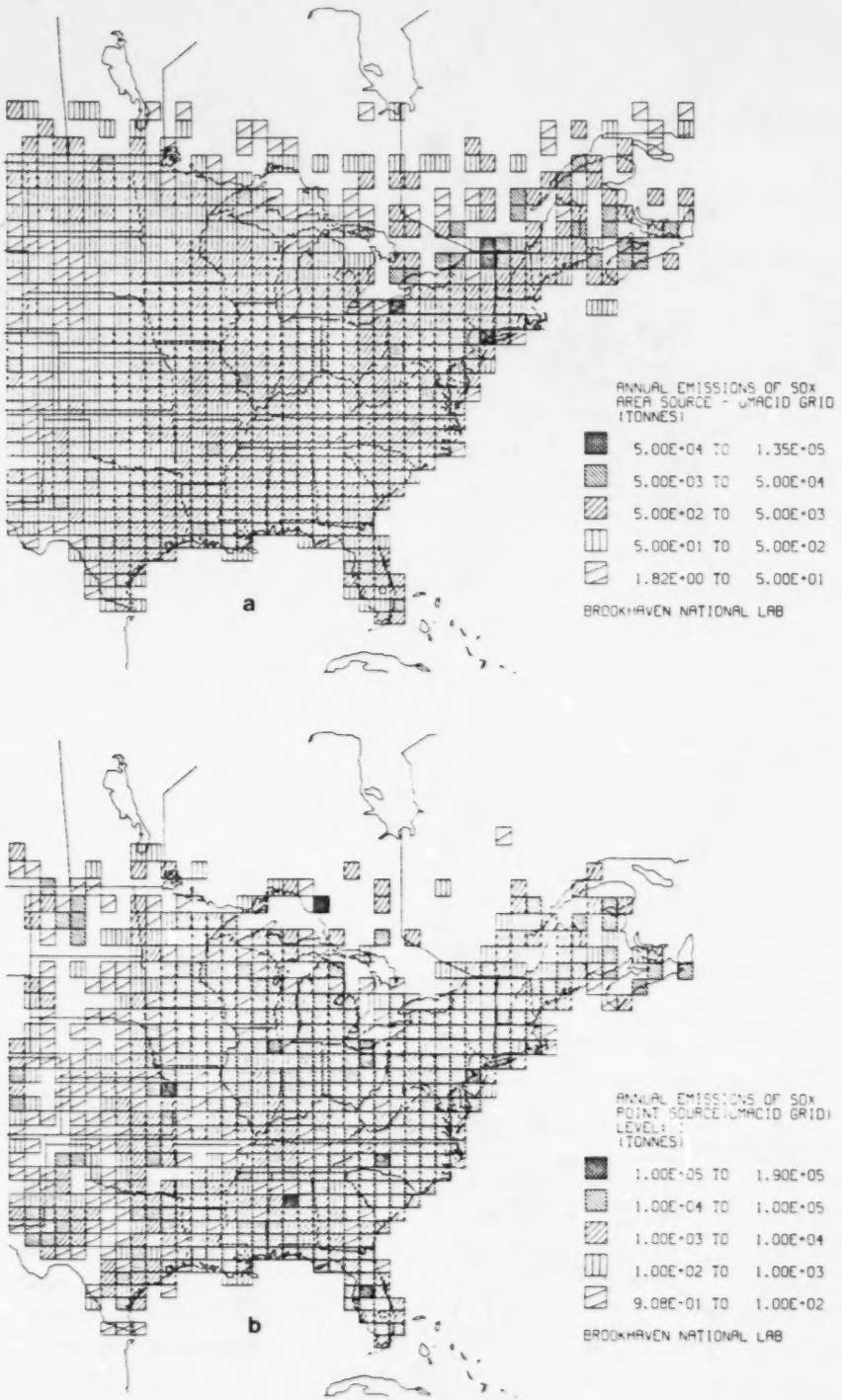
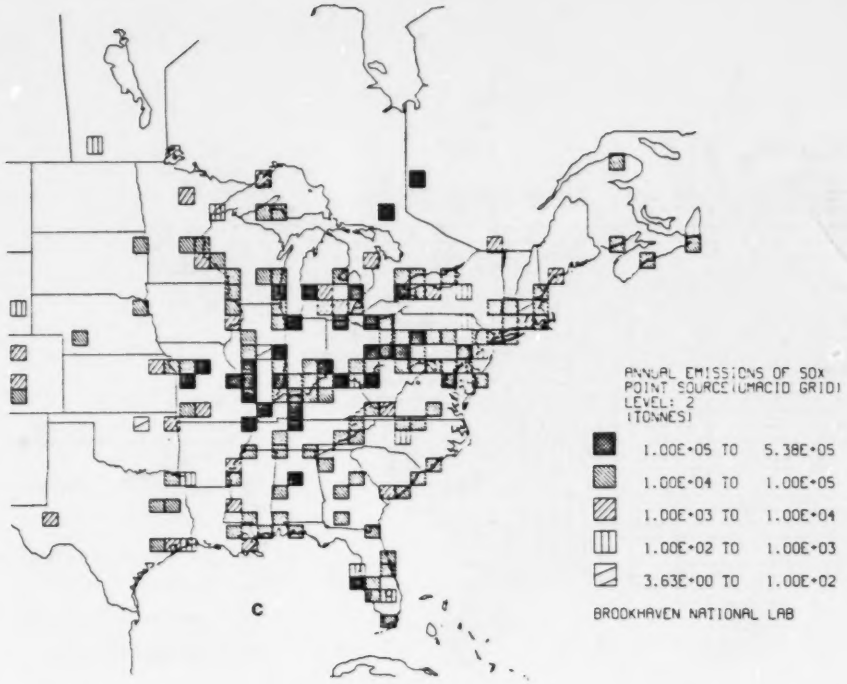

a
ANNUAL EMISSIONS OF SOx
AREA SOURCE - UMACID GRID
(TONNES)
5.00E+04 TO 1.35E+05
5.00E+03 TO 5.00E+04
5.00E+02 TO 5.00E+03
5.00E+01 TO 5.00E+02
1.82E+00 TO 5.00E+01
BROOKHAVEN NATIONAL LAB
b
ANNUAL EMISSIONS OF SOx
POINT SOURCE(UMACID GRID)
LEVEL: 1
(TONNES)
1.00E+05 TO 1.90E+05
1.00E+04 TO 1.00E+05
1.00E+03 TO 1.00E+04
1.00E+02 TO 1.00E+03
9.08E-01 TO 1.00E+02
BROOKHAVEN NATIONAL LAB

Fig. 1(a, b).



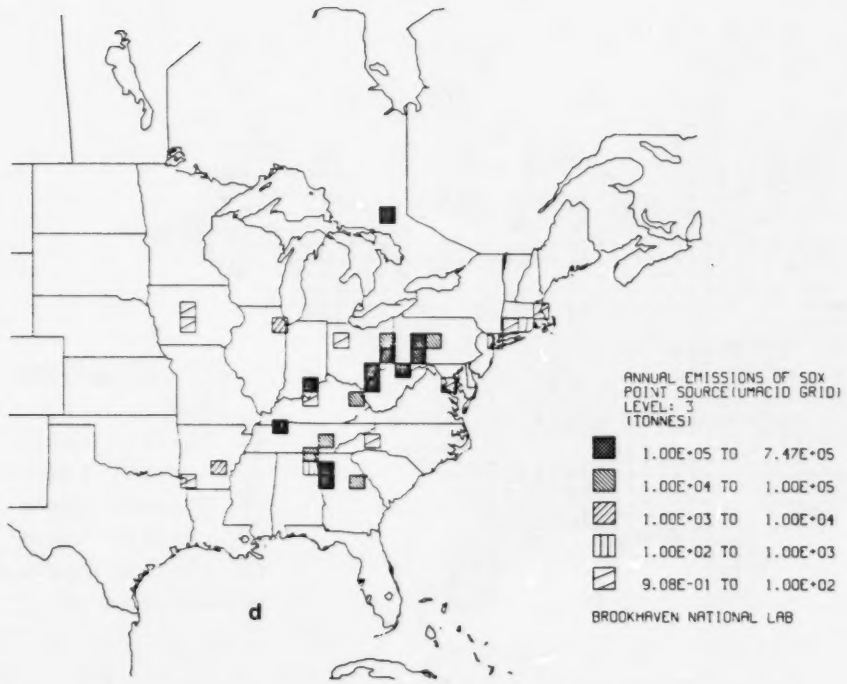


Fig. 1. The 1980 annual SO_2 emission rates (t yr^{-1}) for 80-km grid cells for area sources (a) and point sources with effective stack heights within the (b) 0–200-m, (c) 200–700-m, and (d) greater than 700-m layers.

of their being within 50 km of a region emitting more than 10,000 t yr^{-1} of SO_2, were included to compensate for the otherwise data-void areas. For each season Table 2 indicates the number of sites in each of these overall data quality levels, while Fig. 2 provides the seasonal amounts of sulfur wet deposition at each site.

Since the spatial resolution of the models is of the order of 100 km, colocated or nearly colocated (within 100 km) sites were considered as representing one virtual point whose location was at the midpoint of the group of sites. The seasonal and annual deposition amounts at these points were determined by averaging those amounts with the higher overall data quality levels. If only one site in the group had the higher overall data quality level, only the amount at that site was used to represent the virtual point. Although internetwork data inconsistencies have been observed (De Pena *et al.*, 1980; Vet *et al.*, 1981; Vet and Onloch, 1981), data adjustments were not attempted. Nonetheless, one should be aware that some inconsistencies were found.

5. MODEL EVALUATION APPROACH

5.1. *Spatial patterns*

In response to the need to determine the spatial distribution of acidic deposition, S deposition models have been applied to produce seasonal and annual patterns of S wet and dry deposition. In evaluating the performance of these models, it is therefore essential to assess their ability to replicate the observations not only in probability distribution, but also in spatial pattern. Further, since the observed seasonal amounts of S wet deposition tend to be highest in summer and lowest in winter, the models should be expected to replicate this seasonal distribution, and therefore, be evaluated for seasonal periods.

In an effort to obtain information regarding spatial accuracy, past evaluations have relied on residuals, computed as differences between model predictions and observations at monitoring sites. However, the conclusions drawn from past analyses of residuals alone may be misleading for several reasons. First of all, measurements possess an aggregated error component that models cannot (and should not) be expected to mimic; contributing to this error component are errors associated with precipitation collection efficiency, sample contamination, missing data and sample analysis. The resulting uncertainty should be taken into account in comparisons with predictions. Secondly, since the spatial resolution of the models is of the order of 10^4 km^2, grid cell predictions should not be compared to point measurements without consideration for the spatial variability near the point of measurement. The data at one site may vary considerably from its neighbors because of a local emissions source or the local variability associated with meteorological phenomena. This variation is often large enough to reduce one's confidence in an individual site's ability to accurately represent a larger area.

Table 2. The number of sites in the 1980 ISDME evaluation data set

	Overall data quality level		
	1	2	Total
Winter	24	14	38
Spring	24	22	46
Summer	30	15	45
Autumn	27	15	42

Because of the uncertainty in individual site data, a more reasonable approach for evaluating model performance is one that compares the spatial patterns derived from interpolated predictions and observations. These interpolated patterns retain the general spatial features of the data by smoothing through random measurement error and local spatial variation. Associated with the estimation of spatial patterns, there is an interpolation error which results from these sources of variability. Unfortunately, most of the available interpolation methods are deterministic and provide no estimates of the error associated with their estimated spatial patterns. However, the statistical methods of least squares regression and Kriging, a geostatistical interpolation method (Delhomme, 1978; Clark, 1979), provide the uncertainty estimates for interpolated surfaces.

The regression method oversmooths the data, as illustrated in the ISDME report and yields unreliable error estimates. Kriging, on the other hand, was shown to be more appropriate for model evaluation purposes (Clark *et al.*, 1987). The method is based on G. Matheron's theory of regionalized variables, which was designed to mathematically describe phenomena distributed in space and/or time and provide an appropriate means for solving spatial estimation problems. This technique has been used most extensively in the mining industry but has also been applied in a wide variety of earth sciences research (Barnes, 1980). Analysts of acid precipitation data have found the technique particularly useful. Recent investigations using Kriging include an analysis of spatial trends in hydrogen ion concentration in New York State (Bilonick, 1983), an evaluation of the performance of Kriging in interpolating acid precipitation parameters in the eastern U.S. (Finkelstein, 1983), and a study of temporal/spatial patterns of pH and rainfall (Eynon and Switzer, 1983).

In Kriging, the estimates of error, as well as the patterns produced, are based on the spatial covariance structure of the data. At each interpolation point, weights are computed as a function of an estimated covariance that is assumed to be a function of only the distance between points.

This interpolation method has several desirable features. First, unlike ordinary regression, the interpolated value at a point of measurement equals the

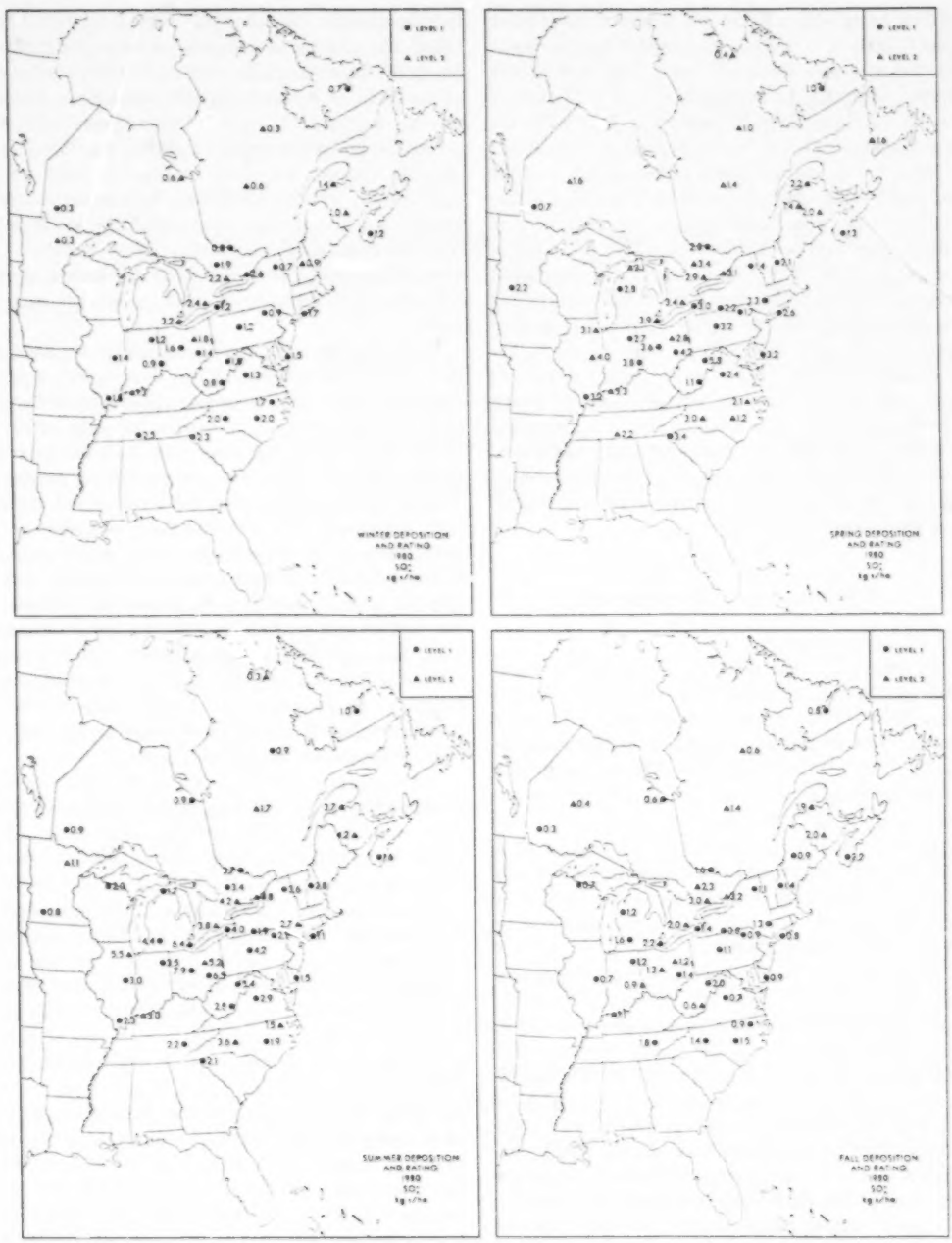


Fig. 2. Seasonal 1980 amounts of sulfur wet deposition (kg S ha^{-1} $season^{-1}$) at the ISDME evaluation sites.

measured value. Secondly, assuming that the semivariogram (i.e. the model estimating the covariance structure of S wet deposition) is correctly estimated, this method yields the lowest variance estimates compared to unbiased methods based on linear functions of the data. Finally, although the variance estimates are biased approximately 15–20%, the bias is not large and its direction is known (i.e. the variances are overestimated). Consequently, pattern comparisons relying on these interpolation variance estimates will be less stringent than they should be (i.e. a wider margin of model error is permitted) (Seilkop, 1986).

5.2. *Model performance measures*

The model evaluation focused on two aspects of spatial pattern replication, namely that of pattern *magnitude* and *position*. Model performance was assessed using the five performance measures identified by Table 3. The primary measure of pattern magnitude was the percentage of area in each of four subregions, depicted in Fig. 3, where the interpolated site predictions are significantly different from the interpolated site observations, as described below. The subregions were defined on the basis of the geographic regions across which the models tended to significantly under/overpredict. The advantage of this measure is that subregions where the models perform best and worst are readily identified by a set of quantities.

This measure was quantified by first applying the simple Kriging method to interpolate at intervals of one-half-degree latitude and longitude the observations and predictions. Predictions were interpolated, since every model did not provide gridded predictions. At each point, the interpolated predictions (P_{ij}) and observations (O_{ij}) were then compared relative to the uncertainty in the patterns. The uncertainty of the annual observed patterns ranged from approximately $\pm 50\%$ across the Midwestern U.S. and southern Ontario to approximately $\pm 100\%$ across Ontario and Quebec.

A confidence interval for the difference between the interpolated predictions and observations was determined from the following expression:

$$P_{ij} - O_{ij} \pm 1.96\,[Var\ P_{ij} + Var\ O_{ij}]^{1/2} \qquad (1)$$

where P_{ij} and O_{ij} are the prediction and the observation, respectively, at point ij and Var is the Kriging variance estimate. Since the deviations of the Kriged values from the true values approximated a normal (Gaussian) distribution (Clark *et al.*, 1987), this is approximately a 95% confidence interval. Failure of the confidence interval to include zero signified that the prediction was significantly different from the observation at this point.

The second measure of pattern magnitude replication focused on the differences in the magnitudes of the interpolated predicted and observed maxima, Δ_{mx}, with consideration of the uncertainty of the interpolations. The value of this measure was defined as

$$\Delta_{mx} = \begin{cases} |P_{mx} - O_{mx}| - 2\sigma_i, & \text{if } |P_{mx} - O_{mx}| > 2\sigma_i \\ 0, & \text{otherwise} \end{cases} \qquad (2)$$

where P_{mx} and O_{mx} are the magnitudes of the interpolated maximum observation and prediction, respectively, and the σ_i is the uncertainty estimate of the maximum observation. Rather than being an indication of concurrence, this measure indicates whether the maximum interpolated prediction lies within the uncertainty of the maximum interpolated observation (when $\Delta_{mx} = 0$), or if not, by how much the maximum interpolated prediction lies outside the uncertainty range.

The replication of pattern position was assessed using three measures. The first measure, the distance between the centroids of the predictions and observations, indicates the degree of concurrence of the two spatial distributions. The location of the centroids of the predictions relative to those of the observations indicates both the distance and the direction toward which the pattern of predictions is shifted.

The centroids are computed as deposition-weighted means of the longitude ($\Sigma w_i x_i$) and the latitude ($\Sigma w_i y_i$), where the weights, w_i, are defined as

$$w_i = \frac{D_i}{\sum_{i=1}^{N} D_i} \qquad (3)$$

and D_i is the amount of S wet deposition at point i. To focus on the high deposition region, only the amounts across the midwestern U.S., the central Appalachian Mountain Range and southern Ontario were considered.

The second measure, namely the difference between the angles of the major axes of the two patterns, quantifies the degree of concurrence in the orientation of the patterns. The major axes were determined from a weighted principal component analysis centered at the centroids, with weights proportional to wet deposition. Both the centroids and major axes based on site observations and predictions showed good agreement with those based on the interpolations.

The last measure defines the differences in the location of the actual and predicted maximum amount of wet deposition. Because of the wide spacing of monitoring sites, the actual locations of the true maximum are uncertain. Hence, the locations are estimated. The distances separating the location of these and the predicted maxima are computed using two different approaches. The first simply computes the distance between the site with the maximum observation and the site with the maximum prediction. The alternate approach uses interpolated observations and predictions, and the uncertainties about them, to define geographical domains where the maximum of each is likely to occur. The shortest distance between the domains is the value of the measure. In general, the results of each approach were comparable.

6. MODEL EVALUATION RESULTS

6.1. *Seasonal distributions*

Reflecting the seasonal variability in the SO_4^{2-} scavenging rates, the observed amounts of S wet deposition peak in the warm months and reach minimal values in the cooler months. Figure 4 illustrates the mean seasonal distribution of S wet deposition for 1980 at the 30 monitoring sites where valid data were

Table 3. Values of the model performance measures for each season of 1980

SUMMER 1380	AES	ASTRAP	MOE	RELMAP	RIVAD	RTM-II	SERTAD	SIMPMOD	STATMOD	TVA	UMACID
PATTERN MAGNITUDE											
Area of Significant Differences (%)[1]											
Northeast	16.9-	1.1-	17.1-	7.5-	26.3-	22.9-	0.4-	29.3-	37.6-	22.3-	27.1
Northwest	4.3-	6.4+	10.0-	5.9+	42.8-	48.0-	3.2+	38.7-	54.8-	16.3-	16.3-
Southwest	14.5-	4.9+	17.1-	32.9+	46.5-	46.9-	4.4-	31.1	29.8-	10.9-	31.1-
Southeast	3.9+	48.7+	3.3-	90.1+	27.0-	26.3-	3.9+	3.3+	11.2-	2.0+	0.7
Composite	9.4-	10.6+	12.4-	23.7+	37.3-	38.6-	2.9	29.9-	39.2-	14.9-	20.0-
Magnitude of Maximum Amount (%)[2]	- 9	+ 1	-34	+85	-60	-59	-18	-32	-42	-14	-30
PATTERN POSITION											
Orientation (degrees)[3]	+ 2.8	+ 8.5	+11.7	+10.3	+ 6.8	+ 6.2	+14.6	+13.4	+19.6	+12.5	- 0.6
Location of Centroid (km)[4]	100	63	51	147	33	54	10	127	92	94	42
Location of Maximum Amount (km)[5]	277	351	84	277	84	84	263	259	164	84	277

AUTUMN 1980	AES	ASTRAP	MOE	RELMAP	RIVAD	RTM-II	SERTAD	SIMPMOD	STATMOD	TVA	UMACID
PATTERN MAGNITUDE											
Area of Significant Differences (%)[1]											
Northeast	1.5+	1.5+	9.8-	0.0	21.4-	0.0	10.5+	0.0	0.0	13.2-	0.0
Northwest	9.5+	14.9+	0.0	4.1-	23.8-	18.8-	14.0+	3.2+	0.2-	0.0	0.0
Southwest	0.0	11.0+	0.0	0.0	0.0	0.0	0.0	18.4+	0.0	2.6+	0.0
Southeast	63.8+	53.9+	0.0	7.9+	0.0	0.0	13.8+	78.9+	14.5+	10.5+	2.6+
Composite	13.1+	16.3+	2.4-	2.8	14.9-	7.6-	10.2+	16.2+	2.1+	5.2	0.4+
Magnitude of Maximum Amount (%)[2]	0	0	0	0	-52	-42	0	0	0	0	0
PATTERN POSITION											
Orientation (degrees)[3]	+ 7.3	+16.8	+18.2	+ 3.3	+12.2	+ 8.8	+13.3	+20.7	+23.8	+21.0	+ 8.7
Location of Centroid (km)[4]	139	65	71-	182	71	97	65	126	147	99	38
Location of Maximum Amount (km)[5]	338	183	164	338	315	135	0	338	285	338	338

1 Significance level approximates 95%; "+" and "−" denote a prevalence of overpredictions and underpredictions, respectively.

2 Percentage difference determined by $[(O_{ms} - P_{ms})/O_{ms}] \times 100$, where P_{ms} is the maximum predicted amount and O_{ms} is the lower/upper value (whichever is closer to P_{ms}) of the uncertainty range about the maximum observed amount.

3 Difference determined by $A_p - A_o$, where A_p and A_o are the angles of the major axes of the patterns of predictions and observations, respectively.

4 Distance between the centroids of the patterns of predictions and observations.

5 Minimum distance between the location of the predicted maximum amount and the isopleth of the lower limit of the uncertainty about the observed maximum amount.

SPRING 1980	AES	ASTRAP	MOE	RELMAP	RIVAD	RTMI-I	SERTAD	SIMPMOD	STMOD	TVA	UMACID
PATTERN MAGNITUDE											
Area of Significant Differences (%)[1]											
Northeast	0.0	0.0	0.0	0.0	0.0	0.4-	0.0	0.0	1.9	0.0	2.3
Northwest	0.5	1.8+	0.0	1.4+	16.1-	38.1-	0.0	0.0	33.0-	0.0	12.9-
Southwest	0.0	0.0	0.0	14.0+	57.5-	64.0-	0.0	0.0	1.8-	0.0	32.0-
Southeast	0.0	20.4+	0.0	59.9+	25.7-	27.6-	0.0	0.0	0.0	0.0	3.3-
Composite	0.2-	3.6+	0.0	11.9+	22.2-	33.3-	0.0	0.0	14.2-	0.0	13.0-
Magnitude of Maximum Amount (%)[2]	0	0	0	+55	-35	-33	0	0	0	0	0
PATTERN POSITION											
Orientation (degrees)[3]	- 2.6	- 3.1	- 0.3	- 0.5	- 2.2	- 2.5	- 1.3	- 1.6	+ 7.2	- 0.3	+ 0.2
Location of Centroid (km)[4]	51	68	66	82	30	25	16	72	94	46	30
Location of Maximum Amount (km)[5]	0	0	0	0	0	0	0	0	0	0	0

WINTER 1980	AES	ASTRAP	MOE	RELMAP	RIVAD	RTMI-I	SERTAD	SIMPMOD	STMOD	TVA	UMACID
PATTERN MAGNITUDE											
Area of Significant Differences (%)[1]											
Northeast	0.0	0.0	0.8-	0.0	0.8-	0.0	0.0	0.0	0.0	0.0	0.0
Northwest	0.2+	11.5+	0.0	5.4-	31.2-	27.6-	0.7+	14.7+	3.4-	0.0	0.9-
Southwest	41.2+	29.4+	0.0	5.3+	4.4-	2.8-	0.0	62.3+	33.3+	4.4+	1.8+
Southeast	85.5+	52.6+	5.3-	50.7+	0.0	0.0	19.7+	98.7+	32.7+	13.2+	9.2+
Composite	20.7+	18.2+	0.8-	10.4+	13.8-	[illegible]	3.0+	32.5+	11.7+	2.8+	2.1+
Magnitude of Maximum Amount (%)[2]	+38	0	0	0	-25	-18	0	+38	0	0	0
PATTERN POSITION											
Orientation (degrees)[3]	+ 5.3	+ 9.7	+13.3	+ 3.7	+10.2	+10.2	+ 9.4	+12.6	+20.1	+14.4	- 0.1
Location of Centroid (km)[4]	148	43	82	144	66	112	42	81	151	32	87
Location of Maximum Amount (km)[5]	238	42	55	238	202	202	197	238	238	101	253

Fig. 3. The four subregions where significant differences between interpolated predictions and observations are identified. Square site symbols denote those 30 sites where observations passed at least level 2 screening criteria for all seasons of 1980.

available for each season. These sites are identified by the small squares in Fig. 3. Approximately 40% and 30% of the annual total occurs during the summer (July, August and September) and spring (April, May and June), respectively, while approximately 15% occurs in both autumn (October, November and December) and winter (January, February and March).

The seasonal distributions of the majority of models resemble the observed distribution. The exceptions are SIMPMOD and STATMOD, two very similar models, which predict considerably greater and lesser amounts for winter and summer, respectively. Winter predictions of AES and UMACID are also relatively high, but both models indicate that the summer amounts are greatest.

6.2. *Pattern magnitude*

Reflecting the daily and seasonal variations in the factors determining the mass of sulfur removed from the atmosphere, spatial patterns of sulfur wet deposition for each of the seasons of 1980 are quite different. The summer pattern (Fig. 5) shows that the greatest amounts of S wet deposition, exceeding 6.0 kg S ha^{-1}, are found across Ohio and western Lake Erie. A ridge of high wet deposition amounts extends from Illinois and Wisconsin to Ohio and continues northeastward into eastern Ontario. Several secondary peaks are apparent to the north and northeast of Lake Ontario and in the vicinity of southern Lake Michigan. Also apparent is the region of minimum amounts along the East Coast of the U.S., which corresponds to the region of minimum summer precipitation amounts.

Unlike those of the summer pattern, the gradients of the spring pattern (Fig. 6) are much less. Maximum amounts, >3.5 kg S ha^{-1}, are located further south than their summer counterparts. The ridge of maximum amounts stretches from eastern Illinois to eastern West Virginia. Areas of minimum deposition are located across eastern North Carolina, the northeastern U.S. and the Maritime provinces.

The autumn pattern (Fig. 7) shows even less spatial variability than those of summer and spring. Maximum amounts, >2.0 kg S ha^{-1}, are located across much of southwestern Ontario, while a pronounced secondary maximum approaching 2.0 kg S ha^{-1} is situated across the Maritime provinces. The low pre-

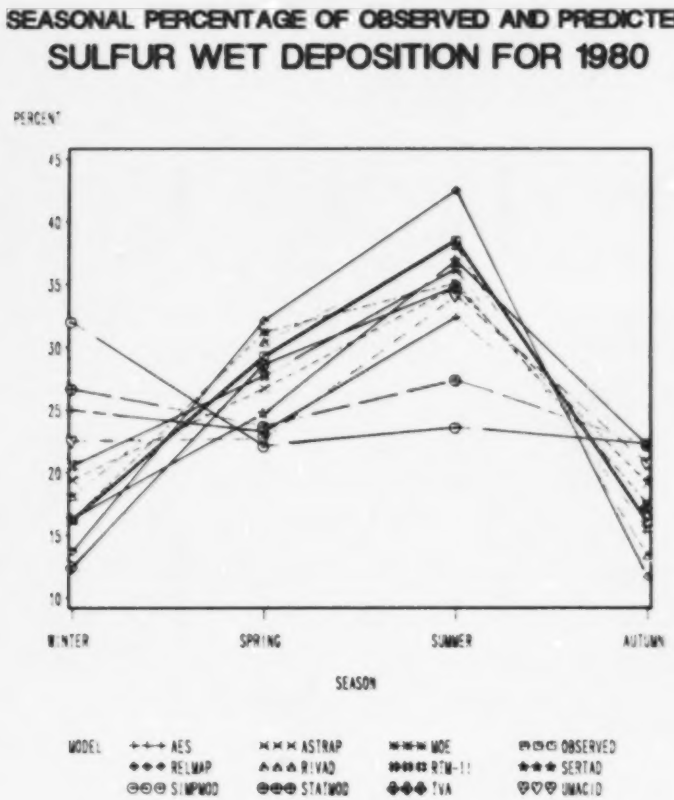


Fig. 4. The mean percentage of annual 1980 sulfur wet deposition observed and predicted for each season at 30 ISDME sites (see Fig. 3 for locations).

cipitation amounts across the high emission region of the U.S. (half of the summer and spring amounts) is a possible explanation for the northward shift of the area of maximum wet deposition amounts. Since data are not available in southern Quebec, the uncertainty of the actual existence of the two distinct areas of maximum amounts is rather high. Nonetheless, it is clear that, unlike the summer and spring patterns, the greatest amounts are found in Canada.

The magnitude of the winter amounts (Fig. 8) resembles that of the autumn amounts, however, unlike the other seasons, the ridge of maximum winter amounts shows less directional dependence. Maximum amounts, exceeding 2.3 kg S ha^{-1}, are found across southwestern Ontario and eastern Michigan. Unlike the other seasons, however, a secondary maximum appears across the southern Appalachian Mountains where the precipitation amounts are 2–3 times greater than those in Canada.

In contrast to the patterns of observed wet deposition, patterns of predicted wet deposition are smoother. That is, the predicted patterns typically exhibit one area of maximum amounts and tend to reduce the spatial variability depicted in the observed patterns. The predicted patterns are similar to each other in that the isopleths resemble sets of nearly concentric ellipses and they tend to exhibit a single area of peak wet deposition amounts in the high emissions region of eastern Ohio, northern West Virginia and western Pennsylvania.

The degree to which the magnitude of the predicted patterns concur with that of the observed patterns is quantified by the per cent of half-degree cells where the predictions are significantly different than the observations according to (2). Table 3 provides the percentages for each of the four subregions as well as the composite region. Based on the analyses of significant differences in pattern magnitude, four types of model behavior, identified in Table 4, are apparent. For three seasons the predictions of SERTAD and TVA, both statistical models, are insignificantly different from the observations virtually everywhere. The other two statistical models, STATMOD and SIMPMOD, behaved inconsistently from one season to another. As illustrated in Fig. 4, these two models tended to significantly underpredict in summer and overpredict in winter.

Reproduced from
best available copy.

Fig. 5. Summer 1980 Kriged analyses of the observations and predictions of sulfur wet deposition (kg S ha^{-1} season^{-1}).

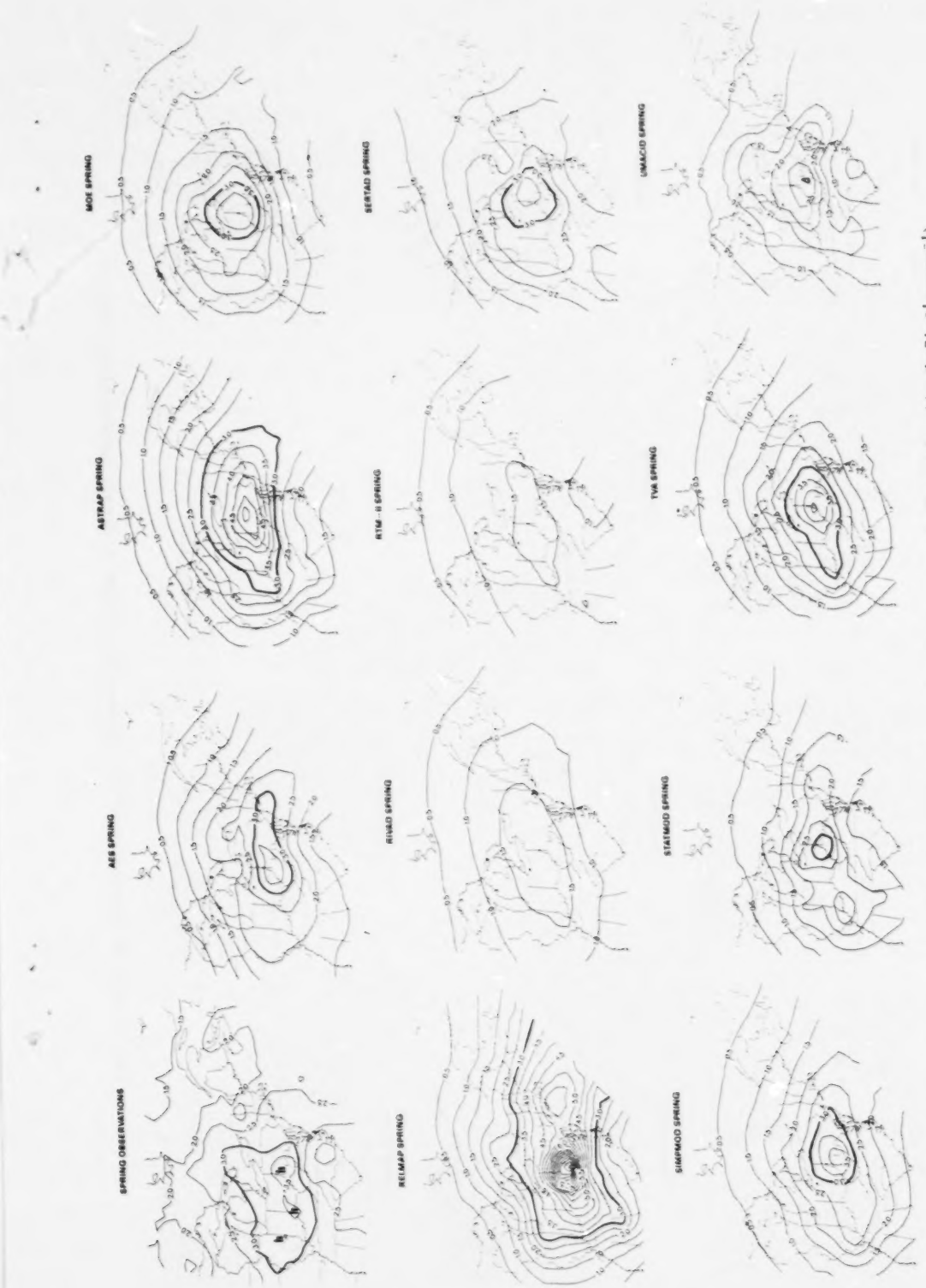


Fig. 6. Spring 1980 Kriged analyses of the observations and predictions of sulfur wet deposition (kg S ha^{-1} season^{-1}).

Reproduced from
best available copy.

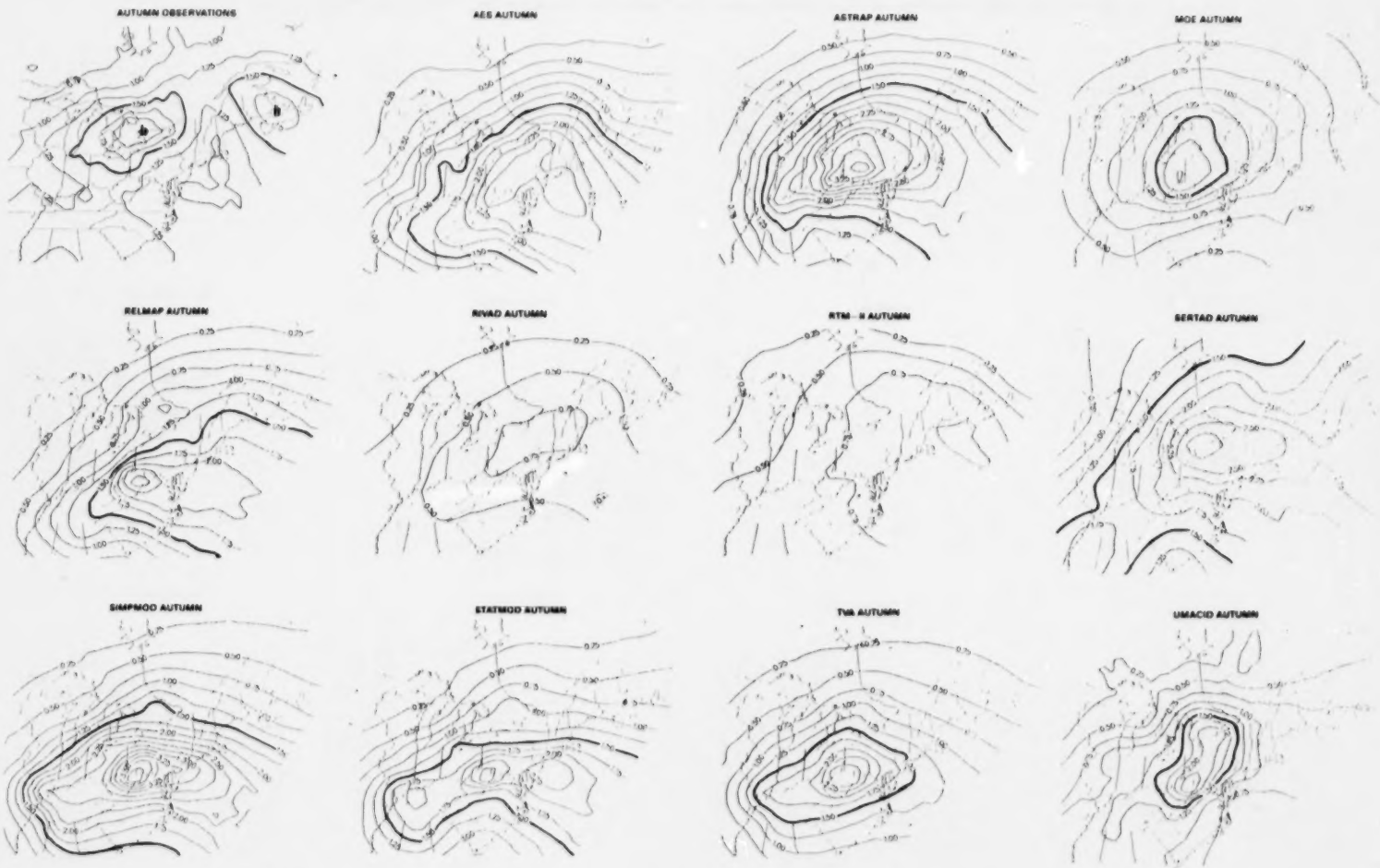


Fig. 7. Autumn 1980 Kriged analyses of the observations and predictions of sulfur wet deposition (kg S ha^{-1} season^{-1}).

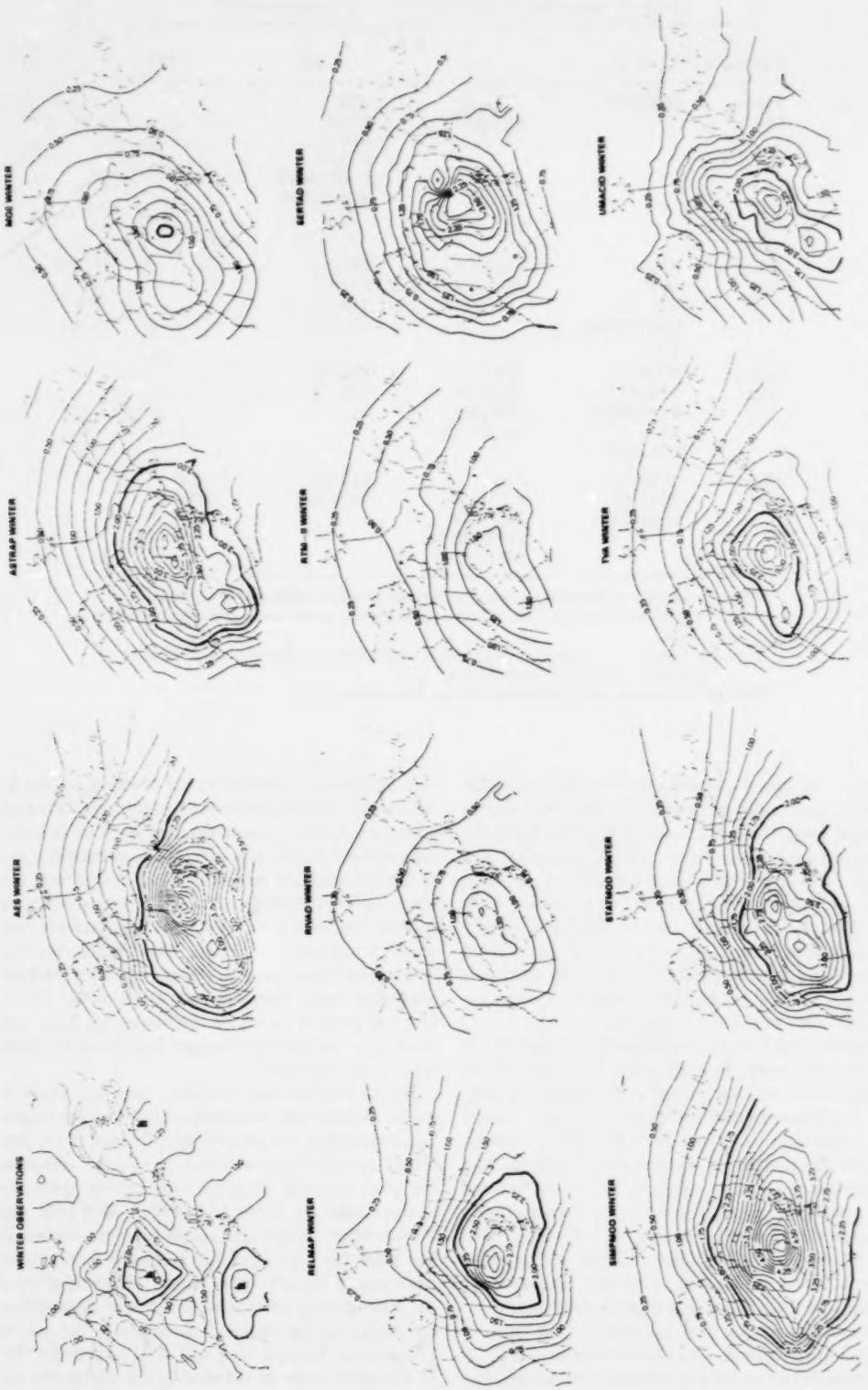


Fig. 8. Winter 1980 Kriged analyses of the observations and predictions of sulfur wet deposition (kg S ha^{-1} season^{-1}).

Table 4. Types of model behavior for the four seasons of 1980

Season	Type* I	II	III	IV
Summer 1980	SERTAD	ASTRAP RELMAP	AES MOE RIVAD RTM-II SIMPMOD STATMOD TVA UMACID	
Spring 1980	AES MOE SERTAD SIMPMOD TVA	ASTRAP RELMAP	RIVAD RTM-II STATMOD UMACID	
Autumn 1980	MOE RELMAP STATMOD TVA UMACID	AES ASTRAP SERTAD SIMPMOD	RIVAD RTM-II	
Winter 1980	SERTAD ASTRAP UMACID	AES ASTRAP RELMAP SIMPMOD STATMOD	RIVAD RTM-II	MOE RIVAD

*Type I: Significant differences nowhere or limited to small widely scattered areas.
Type II: Significant overpredictions across areas to the east and/or south of the Ohio River Valley.
Type III: Significant underpredictions across continuous or sporadic areas from the midwestern U.S. to Nova Scotia.
Type IV: Significant underpredictions across the southern U.S.

Except in spring and autumn when its predictions were insignificantly different from the observations virtually everywhere, MOE significantly underpredicted across portions of the southern U.S. (winter) and across areas from the midwestern U.S. to the Maritime provinces (summer). For every season both ASTRAP and RELMAP tend to significantly overpredict across areas to the east and south of the Ohio River. The extent of this area varies with season. A coding error in the RELMAP transport module explained much of this overprediction.

Likewise, AES tends to significantly overpredict in autumn and winter, but for spring its predictions are insignificantly different virtually everywhere. In summer, it significantly underpredicts across some areas of the midwestern U.S. and the Maritime provinces. UMACID significantly underpredicts summer and spring amounts across portions of the midwestern U.S., but for the other seasons the differences are insignificant virtually everywhere. Finally, both RIVAD and RTM-II, which are similar models, significantly underpredict the amounts across large portions of the composite region for each season.

When interpreting the measure for the area of significant differences, one should bear in mind one potential drawback of this measure: the uncertainty envelope, used to identify those cells where the prediction significantly differs from the observation, can be inflated for those models which exhibit relatively high variances of their predictions compared to the variance of the observations [see (1)]. Consequently in a model evaluation, these models could benefit from this advantage. In the ISDME the RELMAP variances for summer and spring and ASTRAP variances for each season are greater than those of the other models and, as a result, these models would appear to perform somewhat better than they would otherwise. Therefore, the area of significant differences for these two models for the specified seasons represents the lower limit of the estimate.

Unlike the first measure, which compares amounts paired in space, the second measure of pattern magnitude replication, the percentage difference of the predicted and observed maximum amounts, compares the peak seasonal amounts unpaired in space. As demonstrated by Table 3, ASTRAP, SERTAD and TVA predictions of the maximum amounts are within the uncertainty limits of the maximum interpolated observations virtually every season. The predictions of the other models are at least 30% less or greater than the maximum interpolated observations for one or two seasons. With the exception of RIVAD and RTM-II, the predictions of the models are within the uncertainty limits of the maximum interpolated amounts

for at least two seasons. The tendencies of the models to under/overpredict the maximum amounts are similar to those discussed earlier for the first measure.

6.3. *Pattern position*

The orientation of the major axis, a useful measure of pattern position, indicates the direction along which the wet deposition amounts tend to be greatest. The direction of the axis is dependent on several factors, two of which relate to the precipitation pattern and the air flow patterns. The orientation of the major axis of the observed seasonal patterns varies from 30 to 55 degrees, and approximates a southwest-to-northeast direction.

As apparent from Fig. 9 the orientations of the predicted patterns do not vary with season as much as those for the observed patterns. Those deterministic models using actual surface and upper-level wind measurements to determine transport vectors for discrete periods (i.e. AES, RELMAP, RIVAD, RTM-II and UMACID) appear to best replicate the seasonal changes in the orientations of the observed patterns. In fact, with the exception of autumn, UMACID duplicates the orientations for each season.

On the other hand, the orientations for ASTRAP, SERTAD, SIMPMOD, STATMOD and TVA, which use seasonal mean transport vectors, and MOE, which uses geostrophic vectors determined from surface pressure fields, do not vary seasonally as much. Figure 9 also demonstrates that the orientations for the patterns produced by these six models are (i) rotated (clockwise) more from the observed orientations and (ii) differ the most from the observed orientations.

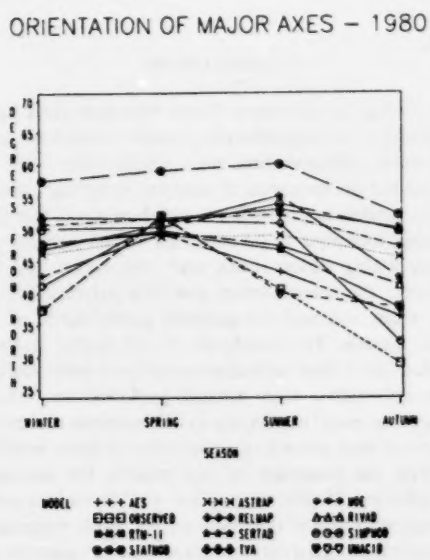


Fig. 9. Seasonal variation of the orientation of the major axes of the seasonal 1980 observations and predictions of sulfur wet deposition.

The fact that the differences for the statistical models are greatest could be a consequence of the methods in which the transport winds are determind. These models use wind roses, or frequency distributions of wind speed and direction, to represent the long-term transport vectors for each simulation period. An underlying assumption of this approach, which may not be valid for all cases, is that the boundary layer flow patterns during precipitation episodes is insignificantly different from the long-term mean flow patterns for the entire simulation period. During the summer, the long-term mean boundary layer flow across the northeastern U.S. tends to be from the west. However, during prefrontal precipitation events when the boundary layer flow tends to be from the southwest, the statistical models would tend to transport and scavenge S to the east of the major source regions, while the deterministic models, which use actual wind measurements, would tend to transport and scavenge S more toward the northeast.

The second measure of pattern position replication, the distance between the centroids of the predicted and observed seasonal patterns, quantifies the spatial shift of the predicted pattern relative to the observed pattern. Despite the fact that the locations of the maximum observed amounts shift seasonally, the locations of the centroid of the observed patterns exhibit small seasonal changes (Fig. 10). Moreover, they tend to congregate near the northern panhandle of West Virginia. The figure also demonstrates that the orientations between the observed and predicted centroids generally do not change appreciably, although the distances do. For instance, the MOE centroids, unlike the others, are always to the north of the observed centroids. Therefore, relative to the observed patterns, the MOE patterns are shifted to the north. In contrast, the other centroids indicate that the patterns of the other models tend to be shifted to the south, southeast or east.

SERTAD appears to best replicate the centroids of the observations. This could be a result of its aforementioned approach in defining its appreciable background contribution.

As Table 3 indicates, the distances between the centroids of the predicted and observed patterns generally do not exceed 150 km. This distance is slightly greater than the spatial resolution of the models. This table also indicates that the locations of the observed and predicted centroids are generally closest in spring (within 85 km) when the locations of the observed and predicted maxima are closest.

Unlike the other two measures of pattern position replication, the third measure focuses on the locations of the interpolated seasonal observed and predicted maxima. The range of distances between these locations (Table 3) changes appreciably from season to season. For each model the distances between the two locations are within the uncertainty limits in spring, and also for SERTAD in the autumn. In summer, however, the predicted maxima are located approx-

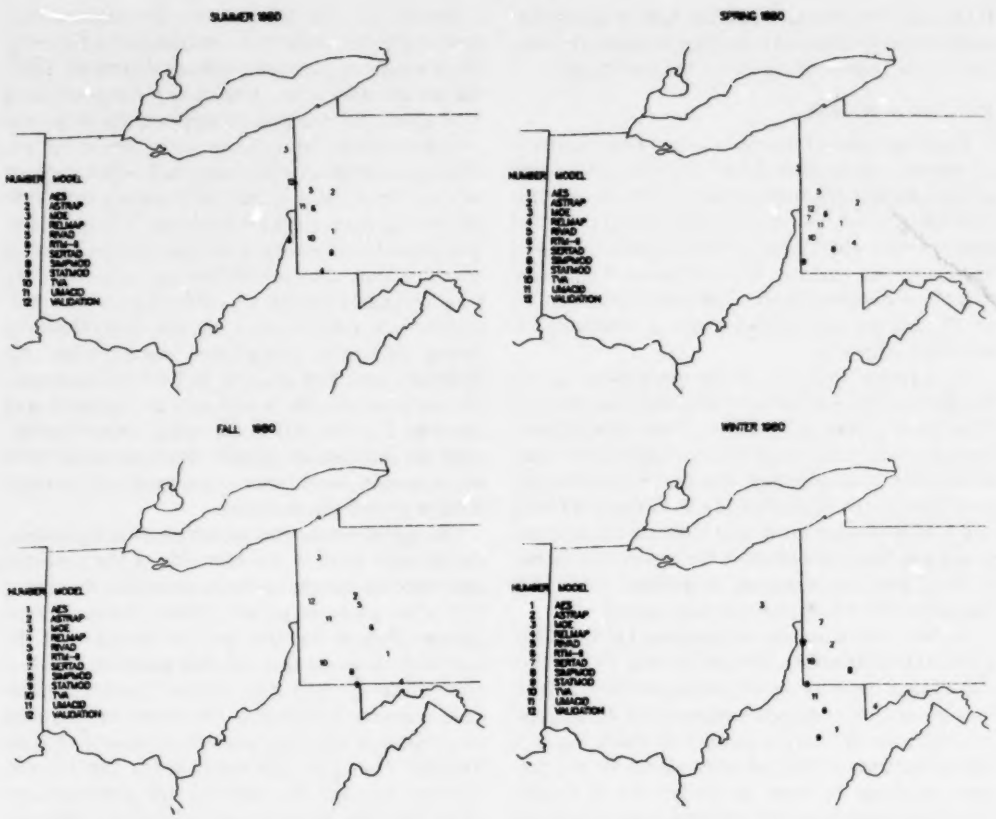


Fig. 10. Seasonal centroids of the interpolated observations and predictions of sulfur wet deposition.

imately 80–350 km to the east or southeast of the location of the observed maximum in central Ohio. Similarly, in winter the predicted maxima are located to the east or southeast of the observed maximum in southwestern Ontario, however, the distances are less, ranging from approximately 40 to 250 km. Unlike those for the other seasons, the locations of the autumn predicted maxima vary to a much greater extent. While the SERTAD location is within the uncertainty limits, both RIVAD and RTM-II maxima are 135–315 km to the east and the maxima for the remaining models are 135–340 km to the south of the observed maximum.

The models have a tendency to predict the maximum seasonal amounts across the Upper Ohio River Valley where SO_2 emissions are greatest (Fig. 1). Since the observed seasonal maxima for 1980 are located in central and southern Ohio during the warm seasons and in southern Ontario during the cold seasons, it appears that the models are insensitive to the factors causing the seasonal shifts. Nonlinear chemistry, which these linear models ignore, could be one of these factors.

7. CONCLUSIONS

Although an extensive North American data base required by a comprehensive model evaluation does not exist, sufficient data are available after 1979 to assess the performance of regional, long-term deposition models in predicting S wet deposition. Using a Kriging technique to (1) construct spatial patterns by interpolating observations and predictions and (2) estimate the measurement and interpolation errors, this study assessed the seasonal performance of 11 such models. The emphasis on (1) spatial pattern replication within estimated uncertainty limits and (2) seasonal, rather than annual, performances establishes this model evaluation as being unique and more rigorous than preceding evaluations of these models.

With the exception of two models, the seasonal distribution of predicted sulfur wet deposition concurred with that of the observations. Both visual and statistical comparisons indicated that the patterns of predictions tended to situate maximum seasonal amounts within the high SO_2 emissions region of the Upper Ohio River Valley, while the maximum ob-

served seasonal amounts migrated from southern and central Ohio in the warm seasons to southern Ontario in the cold seasons. In the spring, every model correctly predicted the location of the observed maximum. For the other seasons, however, the models predicted the location within 170–350 km. Similarly, the centroids across the region of greatest deposition generally indicated an eastward and southward shift in the seasonal patterns of the predictions. However, the pattern shifts are virtually always < 150 km, comparable to the model resolution. These discrepancies could in part be a consequence of the modeling assumption of linear chemistry, which would dictate a linear relationship between emissions and wet deposition amounts.

Also relating to pattern position, the patterns constructed from those models employing a statistical approach to define seasonal mean transport vectors tended to orient their major axes further clockwise, or more toward an east–west direction, relative to the axes of the observed seasonal patterns. The major axes for the predictions of the remaining models were also rotated clockwise, but not significantly. In addition, for these latter models, the seasonal variations in the axis orientations resembled those of the observations. This indicates that those models basing their transport vectors on actual wind measurements for discrete periods were better at replicating the orientation aspect of pattern position.

When data uncertainty limits are considered in the comparison of pattern magnitudes, the seasonal predictions of the Lagrangian model MOE and the statistical models SERTAD and TVA were insignificantly different from the seasonal observations across at least 85% of the evaluation region. However, SERTAD owes much of its success to an imposed background contribution. Thus, this evaluation has demonstrated that a statistical model can perform as well as a Lagrangian model. For unknown reasons, the seasonal performances of the other two statistical models, SIMPMOD and STATMOD, which were very similar to each other, were inconsistent.

AES, ASTRAP and UMACID performed relatively well for two or three seasons, while the seasonal performances of RIVAD and RTM-II were poor due to their tendency to underpredict sulfur wet deposition nearly everywhere for each season. The scavenging rates used by these two models, defined as the ratio of the pollutant concentration in aqueous phase to that in air, appear to be too low. Unlike RIVAD and RTM-II, the performances of the remaining six models varied seasonally for various reasons. One possible reason related to the temporal resolution of the precipitation data used by the models. Both MOE and UMACID use hourly precipitation rates, while the other nonstatistical models, which tend to significantly overpredict, use 6-, 12- or 24-h rates. If UMACID predictions included a uniform background contribution comparable to that of MOE, it would have performed better for the two seasons when it significantly underpredicted across portions of a few subregions.

The results of this model evaluation have highlighted the behavior traits of 11 regional, long-term deposition models. Knowledge of the model strengths and weaknesses benefits both the users and the developers of the models. These results also serve as a benchmark with which subsequent improved versions of the models can be compared. Further evaluation of these models using a more comprehensive data base is encouraged.

Acknowledgements—The authors wish to acknowledge the efforts of all of those who contributed to this evaluation study, especially the modelers and their colleagues who have provided model results, namely Jack Shannon, Argonne National Laboratory; Barbara Ley, Ontario Ministry of the Environment; Doug Latimer, Systems Applications Incorporated; Christopher Fung, Concord Scientific Corporation; Jonathan Pleim, Environmental Research and Technology Incorporated; William B. Norris, Tennessee Valley Authority; and Perry Samson, University of Michigan. The following people have also contributed significantly to this study: Richard Aylesworth, Systems and Informatics Directorate, A. Sirois, Atmospheric Environment Service, Carmen Benkovitz, Veronica Evans and Marty Leach, Brookhaven National Laboratory; Jerry Keeler, University of Michigan; Anthony Olsen, Battelle, Pacific Northwest Laboratory; Dale Coventry, Brian Eder and Peter Finkelstein, U.S. Environmental Protection Agency; Debbie Palka, Analytical Sciences, Incorporated. The authors express their appreciation of the technical comments of Bernard Fisher, Central Electricity Research Laboratories of the U.K., and Dan McNaughton, TRC Environmental Consultants.

REFERENCES

Barnes M. G. (1980) The use of Kriging in estimating the spatial distribution of radionuclides and other spatial phenomena. TRANSTAT (Statistics for Environmental Studies) #13. Battelle Memorial Institute, Richland, WA.

Barrie L. A. and Hales J. M. (1984) The spatial distribution of precipitation acidity and major ion wet deposition in North America during 1980. *Tellus* **36B**, 333–355.

Bhumralkar C. M., Mancuso R. L., Wolf D. E., Nitz K. C. and Johnson W. B. (1981) ENAMAP-1 long-term SO2 and sulfate pollution model: further applications to eastern North America. Final Report, EPA-600/7-81-102, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Bilonick R. (1983) Risk qualified maps of hydrogen ion concentration for the New York State area for 1966–1978. *Atmospheric Environment* **17**, 2513–2524.

Clark I. (1979) *Practical Geostatistics*. Applied Science Publishers, London.

Clark T. L. (1985) Sulfur deposition modeling with the NAPAP emissions inventory. *Proceedings: First Annual Acid Deposition Emissions Inventory Symposium*, EPA-600/9-85-015, Research Triangle Park, N.C.

Clark T. L., Coventry D. H., Eder B. K., Benkovitz C. M., Evans V. A., Volder E. C. and Olson M. P. (1985) International Sulfur Deposition Model Evaluation–1980 standard model input data set, Paper 85-5.3, 78th Annual Meeting of the Air Pollution Control Association, Detroit, MI, 16–21 June.

Clark T. L., Dennis R. L., Palka D., Olson M. P. and Voldner E. C. (1986) Effects of gridded precipitation bias on wet

deposition predictions. Paper 86-29.5, 79th Annual Meeting of the Air Pollution Control Association, Minneapolis, MN, 22–27 June.

Clark T. L., Dennis R. L. and Seilkop S. K. (1989) Reexamination of interim estimates of annual sulfur dry deposition across the eastern United States, EPA-600/4-89/026, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Clark T. L., Dennis R. L., Voldner E. C., Olson M. P., Seilkop S. K. and Alvo M. (1987) International Sulfur Deposition Model Evaluation, EPA/600/3-87/008, U.S. Environmental Protection Agency, Research Triangle Park, NC and ARD-87-1, Atmospheric Environment Service, Downsview, Ontario.

Concord Scientific Corporation (1985) Development of SERTAD: a modeling system for statistical estimates of regional transport and acidic deposition, Part II, Final Report for the Canadian Electrical Association, Downsview, Ontario.

Delhomme J. P. (1978) Kriging in the hydrosciences. *Adv. Wat. Resour.* **1,** 251–266.

De Pena R. G., Pena J. A. and Bowersox V. C. (1980) Precipitation collectors intercomparison study. Department of Meteorology, Pennsylvania State University, State College, PA.

Eliassen A. (1978) The OECD study of long range transport of air pollutants: long range transport modeling. *Atmospheric Environment* **12,** 479–487.

Eynon B. and Switzer P. (1983) The variability of rainfall activity. *Can. J. Stat.* **11,** 11–24.

Finkelstein P. L. (1983) The spatial analysis of acid precipitation data. *J. clim. appl. Met.* **23,** 52–62.

Fox D. G. (1981) Judging air quality model performance–review of the Woods Hole Workshop. *Bull. Amer. Met. Soc.* **62,** 599–602.

Klug W. and Lupkes C. (1985) A comparison between long term interregional air pollution models. NR. 104-025-65, Technische Hochschule, Darmstadt, F.R.G.

Lehmhaus J., Saltbones J. and Eliassen A. (1986) Calculated and observed data for 1980 compared at EMEP measurement stations. EMEP/MSC-W Report 1/86 (Annex), Norwegian Meteorological Institute, Oslo.

McNaughton D. J. (1980) Initial comparison of SURE/MAP3S sulfur oxide observations with long-term regional model predictions. *Atmospheric Environment* **14,** 55–63.

MOI (1982) U.S./Canadian Memorandum of Intent on Transboundary Air Pollution. Atmospheric Science and Analysis Work Group 2 Phase III Final Report.

Norwegian Meteorological Institute (1984) Report of the Meteorological Synthesizing Centre, West (MSC-W) for 1984. Final Report, EMEP/MSC-W Report 1/84, Oslo.

Olsen A. R., Bigelow D. S., Chan W. H., Clark T. L., Lusis M. A., Misra P. K., Vet R. J. and Voldner E. C. (1989) Unified wet deposition summaries for North America: data summary procedures and results for 1984. *Atmospheric Environment* (in press).

Pasquill F. and Smith F. B. (1983) *Atmospheric Diffusion: Study of the Dispersion of Windborne Material from Industrial and Other Sources*, third edition. Ellis Horwood Limited, Chichester.

Perhac R. M. (1978) Sulfate Regional Experiment in northeastern United States: the 'SURE' program. *Atmospheric Environment* **12,** 641–647.

Ruff R. E., Nitz K. C., Ludwig F. L., Bhumralkar C. M., Shannon J. D., Sheih C. M., Lee I. Y., Kumar R. and McNaughton D. J. (1985) Evaluation of three regional air quality models. *Atmospheric Environment* **19,** 1103–1115.

Schiermeier F. A. and Misra P. K. (1983) Evaluation of eight linear regional-scale sulfur models by the Regional Modeling Subgroup of the United States/Canadian Memorandum of Intent Work Group 2. Transactions of the Air Pollution Association Specialty Conference on the Meteorology of Acid Deposition, pp. 330–345.

Seilkop S. K. (1986) Statistical interpolation procedures for mapping ecological data. In *Research Data Management in the Ecological Sciences* (edited by William K. Michener). University of South Carolina Press, Columbia, SC.

Shannon J. D. (1985) User's guide for the Advanced Statistical Trajectory Regional Air Pollution (ASTRAP) model, EPA/600/8-85/016, U.E. Environmental Protection Agency, Research Triangle Park, NC.

Shannon J. D., Kleinman L. K. and Benkovitz C. M. (1982) Intercomparison of MAP3S models of long-range transport and deposition. *Proceedings of the Third Joint Conference on Applications of Air Pollution Meteorology*, 12–15 January San Antonio, TX, pp. 29–32.

Stewart D. A., Morris R. E. and Liu M. K. (1983a) Evaluation of an episodic regional transport model for a multi-day sulfate episode. *Atmospheric Environment* **17,** 1225–1252.

Stewart D. A., Morris R. E. and Liu M. K. (1983b) Evaluation of long-term regional transport models of interest to the National Park Service. Final Report SYSAPP 83/215, Systems Application Incorporated, San Rafael, CA.

Unified Deposition Data Base Committee (1985) A unified wet deposition data base for eastern North America: data screening, calculation procedures, and results for sulphates and nitrates (1980). Final Report, Prepared for the Canadian Federal-Provincial Research and Monitoring Coordinating Committee 79 pp.

U. S. Environmental Protection Agency (1987) Rocky Mountain acid deposition model assessment: review of existing mesoscale models for use in complex terrain. Final Report, EPA/600/3-87/013, Research Triangle Park, NC.

Vet R. J., Chan W. H. and Lusis M. A. (1981) An intercomparison study of three precipitation sampling networks in Ontario—APIOS, CANSAP and GLPN, Ontario Ministry of the Environment. Final Report ARB-002-81-ARSP, Toronto, Ontario.

Vet R. J. and Onlock S. G. (1981) Detailed analyses of the Pennsylvania State University intercomparison of precipitation collectors data. Concord Scientific Corporation Final Report #52-80, Toronto, Ontario.

Voldner E. C., Olson M. P., Oikawa K. and Loiselle M. (1981) Comparison between measured and computed concentrations of sulphur compounds in eastern North America. *J. geophys. Res.* **86 C6,** 5339–5346.

Voldner E. C. and Shannon J. D. (1983) Evaluation of predicted wet deposition fields of sulfur in eastern Canada. Transactions of the Air Pollution Control Association Specialty Conference on the Meteorology of Acid Deposition, 16–19 October, Hartford, CT, pp. 387–399.

Voldner E. C., Sirois A. and Clark T. L. (1984) Data screening and calculation procedures for the North American precipitation chemistry data to be used in the International Sulfur Deposition Model Evaluation. *Proceedings of the Air Pollution Control Association/American Society for Quality Control Specialty Conference on Quality Assurance in Air Pollution Measurements*, 14–18 October, Boulder, CO.

Watson C. R. and Olsen A. R. (1984) Acid Deposition System (ADS) for statistical reporting: system design and user's code manual. Final Report, EPA-600/8-84-023, U.S. Environmental Protection Agency, Research Triangle Park, NC.